AAAI Technical Report WS-13-04
When the City Meets the Citizen 2

Observation of Human Dynamics in Cote d'Ivoire Through D4D Call Detail Records

Ken Wakita
Dept. Mathematical and Computing Science,
Tokyo Institute of Technology
JST/CREST

Ryo Kawasaki and **Masanori Takami**
Dept. Mathematical and Computing Science,
Tokyo Institute of Technology

Abstract

The article outlines how we have identified the cities, city population, strength of social ties among cities, urban mobility in the largest city of Abidjan, and locations of residential and work areas. To deal with otherwise inadequate information contained in the D4D call detail records we introduced some daring assumptions and invented approximation techniques. Although further investigation and assessment are required, these techniques let the dataset illustrate the societies found in the nation.

Keywords:

Population estimation, Human dynamics, Urban mobility, Spatio-temporal reasoning

Introduction

Cote d'Ivoire is a country in West coast of Africa whose population is estimated to be about 20 million. It is a multiethnic society that is formed from six major ethnic groups that differ linguistically, culturally, and religiously. Growth of immigrants during the 70's and 80's added ethnic variety to the country, and it is reported that the number of minor ethnic groups counts to about 60. The country is respected for its higher economic growth rates in the 80's and 90's. However, it suffered from domestic conflict that caused Ivorian civil war, which brought a decade of political troublous situation.

The research was at first motivated by our interest in the structure of the multi-linguistic, multiethnic culture, and started by attempting to identify ethnic groups and weak ties that connect different ethnic groups, expecting the social network clustering techniques to give us insight into the social structure of the country. Later, we were more interested in the structure and relations of cities and urban mobility.

The article briefly reports our estimation and approximation techniques that we have used to compliment the otherwise inadequate D4D dataset(D.Berry 2011; Boyd and Crawford 2011; Eagle 2009; Latour 2010). The techniques have discovered boundaries of cities, estimated population of the cities which are in line with the population statistics, and urban mobility that is used to segment the largest city into residential and non-residential areas. The work presents that coarse-grain, station-to-station, heavily anonymized call detail records can illustrate the social dynamic of the nation in contrast to previous work which successfully identified livehoods from fine grained personal records(Justin Cranshaw 2012).

Identification of Major Cities

This section describes the techniques that we have used to estimate city boundaries, population of the area covered by each antenna, and social tie strength between cities. These techniques are based on inter-antenna geographical proximity as well as the amount of communication between antennas.

Firstly we have tried to estimate the strength of social ties ($w_{i,j}$) between coverage areas of two antennas using the following formula:

$$w_{i,j} = c_{i,j}/(p_i \times p_j)$$

where $c_{i,j}$ is the number of calls made between the antennas and p_i is the estimated population of the area. It reflects our assumption that the number of calls made from an antenna should be proportional to the population of its coverage area.

As for the special case, where $i = j$, we have the following intra-antenna estimation:

$$w_i = \sqrt{w_{i,i}} = \sqrt{c_{i,i}}/p_i$$

In this formula, the new parameter w_i can be interpreted as digital fluency (or digital addictedness) of the people in the coverage area i.

We assumed that w_i's are equal nation-wide and therefore $w_j = w_i$ for every pairs of i and j.

Figure 1 presents the result of this analysis. To avoid visual clutter, we have trimmed the edges with weaker ties with regard to $w_{i,j}$. We have also reduced the number of nodes by segmenting the whole map into 30×30 grids and replacing antennas in the same grid by their barycenter. Grids are depicted by small circles, and strong ties between grids, namely pairs of grids whose accumulated $w_{i,j}$ are high, are connected by lines. In this figure, we can find many long-distance lines span out from grids in large cities such as Abidjan and Bouakè to grids all over the country. It should also be noted that neighboring grids are connected with short lines. When we further remove inter-grid edges

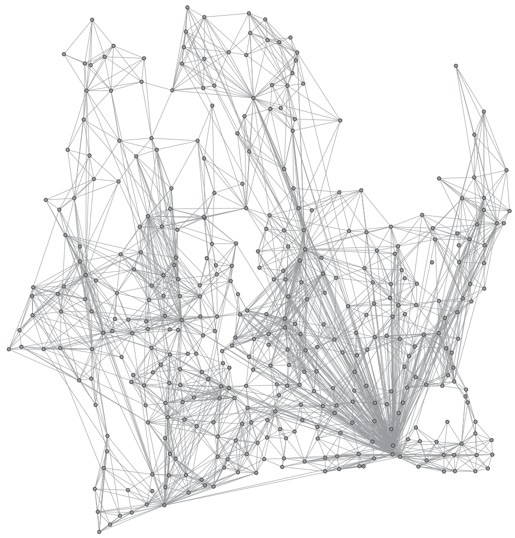


Figure 1: Strength of social ties between antennas

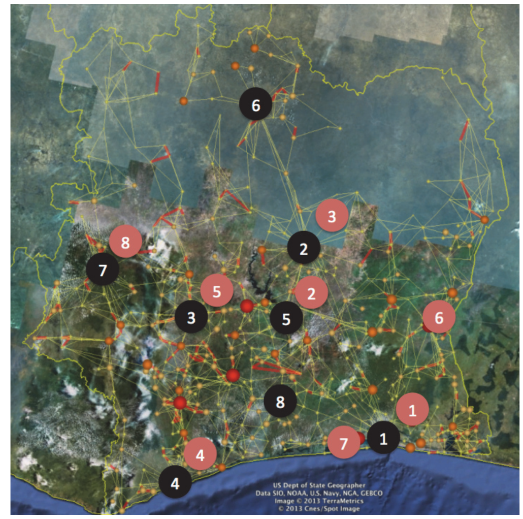


Figure 2: Identification of major cities. Eight major cities are in black and the estimated cities are in red.

by raising the threshold of tie-strength, most of the long-distance edges disappear and we can observe paths forming short lines, revealing a highway net.

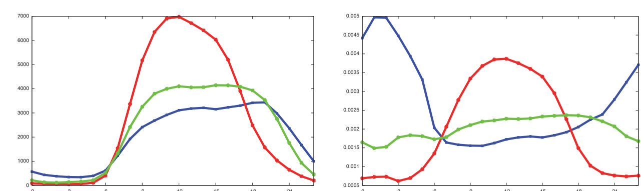
Secondly, we attempted to identify the city boundaries by means of geographical proximity between antennas. However, this technique, being too aggressive in merging neighboring cities, produces poor result. Therefore, we then considered the strength of ties between antennas. The process of finding cities is essentially segmentation of antennas by calculating the closures of geographical proximity and the strength of social ties.

As we have already estimated the population in the coverage areas of antennas p_i, we can easily accumulate them and estimate the population of the city. Figure 2 presents the result. The black circles are locations of eight largest cities documented in (Canty and LLC 2011) and the red circles denote the locations of eight largest estimated cities that are found by our analysis. The circles are labeled by numbers in the descending order of population. Our method seems to over-estimates the population of Dabou (Red-7), which is ranked at 14th position in (Canty and LLC 2011). Korhogo in north (Black-6) which is the 6th largest city, was ranked 11th in our method. Nevertheless, our method gives surprisingly good estimate of the population despite difficulties such as the young generation phone users not found in the dataset, economic discrepancies, differences of education level, divergence of antenna density, and etc.

Urban Human Dynamics in Cote d'Ivoire

The previous section has performed structural analysis of Cote d'Ivoire to identify its major cities, whereas, this section deals with temporal population changes near the antenna sites and tries to segment them into classes such as residential areas and business districts. We also show that the analysis result gives an insight into character and historical background of each city.

For this analysis, we used the second D4D dataset (Set 2), which gives trajectory records of 50,000 sampled mobile users. We have done statistical analysis of the dataset to estimate the temporal population of the coverage area for each antenna, following the technique proposed by (Terada, Nagata, and Kobayashi 2012) for call detail records. In contrast to this previous work which took advantage of their fine-grained records including per-person records of radio beacons, the analysis of D4D Set2 is much more difficult because it is the record of a small sampled mobile users and tracks only the phone calls and text message exchanges, and does not give the trajectory when one is not communicating over the cellar network. To deal with this problem, we have marked the mobile user as *missing* when we can not find one's records in the trajectory database.

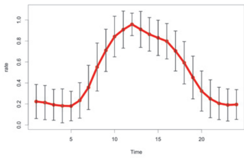


(a) Estimated number of phone users

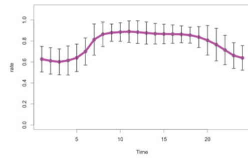
(b) Estimated population distribution over Abidjan

Figure 3: Estimated temporal population patterns three antennas sites in Abidjan #1022 (Blue), #916 (Red), and #772 (Green).

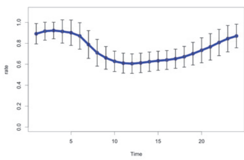
The chart of Figure 3-(a) is the illustration of the result of our modified analysis, that shows temporal estimation of the number of active phone users, where the coordinates stand for hours ($\in [0:23]$) and estimation of the number of active phone users. Three antennas were taken from the largest city of Abidjan, the red line from the city center, the blue from northeast border, and the green in between of them. These lines present average estimated population over the period



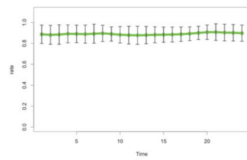
(a) Type I bussiness pattern, 62 antennas



(b) Type II business pattern, 138 antennas



(c) Residential pattern, 165 antennas



(d) Other pattern (including traffics?), 222 antennas

Figure 4: Centroids of the result of k-means clustering for 1,214 antennas gives typical temporal population patterns.

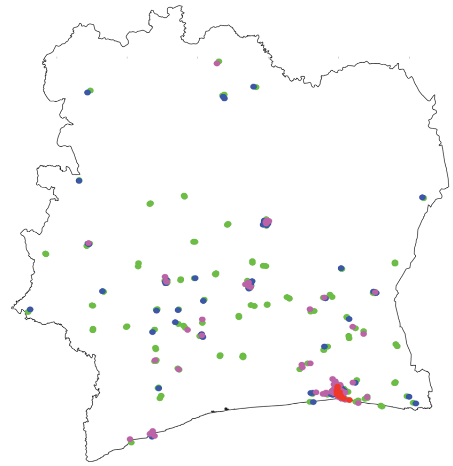
of the dataset (two weeks). We can see that during the night, populations of these antennas are almost emptied since very little communication takes place during the night.

Figure 3-(b) illustrates estimation of populations of the antenna sites. This estimation has been conducted by introduction of two assumptions. Firstly we assumed that population mobility is limited within each city and neglected inter-city mobility. Secondly, the ratio between active phone users in the antenna site relative to the population there is assumed to be equal for all the antenna sites. Then for each time slot, the estimated number of phone users of an antenna site (Figure 3) was normalized against the number of collective phone users over the city to estimate the population distribution of the city. This technique has been applied to all the antenna sites contained in the cities identified in the last section.

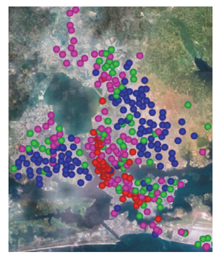
Figure 3-(b) illustrates interesting characteristics of the temporal population distribution of the three antenna sites. The population distribution of antenna site #1022 (blue line) is higher in the night and lower in the daytime (residential areas), whereas antenna site #916 exhibits the opposite pattern (non-residential that include business districts including commercial and industrial uses). We have seen similar temporal patterns in many other antenna sites.

These results suggest the possibility to classify antenna sites according to the temporal patterns of their population distribution. For this purpose, we applied k-means clustering ($k = 4$). Each antenna site has been modeled by 24-dimensional normalized vector that represents temporal population distribution over 24 hours. The distance is measured by Euclidean distance.

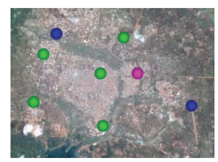
Figure 4 is an error bar chart that illustrates the centroids of the four clusters obtained from Abidjan by this analysis. Chart (a) presents bustle during the daytime and chart (b) tends to be so. They represent temporal population patterns



(a) Nation-wide



(b) Abidjan



(c) Abengourou

(d) San-Pédro

Figure 5: Classification of antennas. Red and magenta: Type I and II business district, blue: residential areas, green: others

of business districts that include industrial and commercial areas. Chart (c), on the other hands, host population in the night and is represents temporal population patterns of residential areas.

Figure 5 illustrates the locations of residential areas and business districts in the nation. The color scheme follows that of Figure 3. It is observable that magenta and blue dots are scattered all over the nation but red dots are found only in Abidjan. It indicates pecuriarity of the highly modernized city in comparison to other smaller cities. In the following, we are going to visit some interesting cities from the map.

Figure 5-(a) is a scale-up of the map around Abidjan. It is by far the largest city in Cote d'Ivoire that has 3.6 million population. In the south of the map, we can find the Marcory island which is segmented into the commercial region (red) in the southwest and residential towns (blue) in the northwest. In the center of the image, we can see a red region of Le Plateau which is the political and economical center of the nation. To its northeast, spans a large blue region that corresponds to a high-class residential area of Cocody. In reality, Le Plateau and Cocody are separated by a high-class commercial area which are depicted by magenta dots. To the northwest of the center of Abidjan is a new industrial area of Yopougon (magenta), to its south is a large residential area (blue), which are in line with recent development in

this area.

Visualization of most middle-scale cities do not present such a clear segmentation as Abidjan. Red dots occur only in Abidjan. For example, k-means clustering for the city of Abengorou (see Figure 5-(b)) presents magenta antennas (type II business district) in the center surrounded by blue (residential areas) antennas. More detail investigation showed that the green antennas have trend toward business districts. There are a few hypothesis that explains this contrast between Abidjan and other cities. Firstly, because Abidjan is by far the largest city (a third of the antennas are installed in Abidjan!) in this nation, our analysis is biased to the characteristics of this super-large city and was insufficiently characterize others. Secondly, it would be the case that Adbijan and other cities are in different stage of industrialization and development and local bussiness is run in the neighbor of residential areas.

During our observation, we found a different pattern in the fourth most populated city, San-Pédro (see Figure 5-(c)). As we have explained, most city has its business district in this center but San-Pédro has one to the east border of the city. We believe it is due primarily from the city being the largest port of shipment of cacao beans and have port next to the eastern gulf of the city. The temporal population distribution for magenta dots have peaks in the morning and suggest the transfer of cacao from other places.

We also found a strange, isolated antenna which is not contained in any of the identified city. This antenna turned out to exist at a fishing town on the bank of Lac de Buyo (lake Buyo), which hosts seasonal visits of a large group of fishermen from Bali.

Summary

In this work, we proposed spatiotemporal analysis techniques for call detail records that identify the location and border of cities, estimate population of the coverage areas of antennas, and classified the coverage areas into residential and non-residential areas. Our technique is useful to some extent for finding interesting human dynamics in Cote d'Ivoire.

In spite of difficulty that the dataset is highly anonymized and only station-to-station collective data is included, no-contents data provided, the analysis has presented a surprising detail of Cote d'Ivoire and suggests effectiveness of this kind of analysis. The effectiveness of the work is partly supported by wide-spread use of cellular phones in this nation as well as its reliance on cellular network from lack of reliable land-lines. We are currently working on investigation of linguistic/ethnic diversity of Cote d'Ivoire.

Acknowledgment Our study and research activity were performed using mobile communication data made available by France Telecom and Orange Cote d'Ivoire within the D4D Challenge. This research was partly supported by the Japan Science and Technology Agency (JST), the Core Research of Evolutionary Science and Technology (CREST) research project. We are thank to Akira Sato, a deputy director of IDE JETRO, for providing us with information on Cote d'Ivoire. We are also grateful to valuable comments from Tsuyoshi Murata and Xin Liu.

References

Boyd, D., and Crawford, K. 2011. Six provocations for big data. *Computer, 123(1)* 1–17.

Canty, and LLC, A. 2011. Cote d'Ivoire - top 100+ cities by population.

D.Berry. 2011. The computational turn: Thinking about the digital humanities. *Culture Machine* 12:1–22.

Eagle, N. 2009. Engineering a common good: Fair use of aggregated, anonymized behavioral data. *First International Forum on the Application and Management of Personal Electronic Information*.

Justin Cranshaw, Raz Schwartz, J. H. N. S. 2012. The livehoods project: Utilizing social media to understand the dynamics of a city. In *AAAI Publications, Sixth International AAAI Conference on Weblogs and Social Media*.

Latour, B. 2010. *Tardes idea of quantification*. M. Candea (ed.) The social after Gabriel Tarde. Routledge.

Lazer, D.; Pentland, A.; Adamic, L.; Aral, S.; Barabasi, A.; Brewer, D.; and et al., N. C. 2009. *Computational social science*. Science, 323(February). 721–723.

Terada, M.; Nagata, T.; and Kobayashi, M. 2012. Population estimation techniques for the mobile spatio statistics (in Japanese). *NTT DOCOMO Technical Journal* 20(3):11–16.